Model Formulation

This document sets out our delayed death model formulation and statistical model assumptions, explains its implementation, and presents the results.

Background

The aim of this analysis is to describe time course data about the viable cell density of CHO cell cultures that were given cytotoxic treatments that induce cell death through a process called apoptosis. Various single and combinatorial knockout cell lines ($\Delta bak1$, Δbax, and Δbok) were generated with the aim of testing apoptosis resistance. The macroscopic process of death was of interest in this study. A population balance model was generated to describe the proposed phenomena of quiescence, delay until death commitment, and the cell death.

Methods

Population balance model formulation

Time evolution of cell state populations is first described as a system of differential equations for which an analytical solution is determined and used for parameter estimation. We assume that in this scenario viable cells exist in four states:

- R replicative cells, growing at a rate of $\mu R(t)$, where t is the current time and $R(t)$ is the current density of replicative cells
- Q_a growth arrested cells, transferring from a replicative state at a rate of $k_q R(t)$
- Q_c death-committed cells transferring from growth arrest at a rate of $k_q R(t-\tau)$, where $\tau > 0$ represents the delay between growth arrest and death commitment.
- Dead, transferring from death committed at a rate of $k_d Q_c(t)$, where $Q_c(t)$ is the density of death-committed cells at time t.

The system of ordinary differential equations (containing a delay differential equation) can be solved analytically, so that the density at a given time can be found as a function of the parameters μ, τ_D k_q, k_d and the initial density of predicative cells R_0.

For cells which are able to replicate;

$$\frac{dR(t)}{dt} = (\mu - k_q)R(t) = \sigma_\mu R(t)$$

Solving for the interval $0 \leq t$,

$$R(t) = R_0 e^{\sigma_\mu t}$$

For cells which are quiescent, but have not committed to a death transition;

$$\frac{dQ_a(t)}{dt} = k_q(R(t) - R(t-\tau_D))$$

Solving for the interval $0 \leq t < \tau_D$;

$$Q_a^{(1)}(t) = R_0 kq \int_0^{t=t} e^{\sigma_\mu t}\, dt = \frac{k_q R_o}{\sigma_\mu}(e^{\sigma_\mu t} - 1)$$

Now, for the interval $\tau_D \leq t$;

$$Q_a^{(2)}(t) = k_q \int_0^{t=t} R(t)\, dt - k_q \int_0^{t-\tau_D} R(t)\, dt = \frac{k_q R_0}{\sigma_\mu}(e^{\sigma_\mu t} - e^{\sigma_\mu (t-\tau_D)})$$

For cells which have committed to death and are transitioning to dead cells;

$$\frac{dQ_c(t)}{dt} = k_q R(t-\tau_D) - k_d Q_c(t)$$

This is solved as a linear first order ODE, with the form;

$$\frac{dQ_c(t)}{dt} + k_d Q_c(t) = k_q R(t-\tau_D)$$

Using the integrating factor $e^{\int k_d\, dt} = e^{k_d t}$

$$e^{k_d t} Q_c(t) = k_q R_0 \int_0^{x=t-\tau_D} e^{k_d(x+\tau_D)} e^{\sigma_\mu x}\, dx$$

where $x = t - \tau_D$

Yielding,

$$Q_c(t) = \frac{k_q R_0}{\sigma_\mu + k_d}\, (e^{(\sigma_\mu + k_d)(t-\tau_D)} e^{k_d(t-\tau_D)} - e^{k_d(t-\tau_D)})\, e^{-k_d t}$$

Total quiescent cells are taken as $Q(t) = Q_a(t) + Q_c(t)$ and total viable cells are taken as $X(t) = R(t) + Q_a(t) + Q_c(t)$ where,

$$
\begin{aligned}
R(t) &= R_0 e^{\sigma_\mu t} \\
Q_a(t) &= \frac{k_q R_0}{\sigma_\mu}(e^{\sigma_\mu t} - 1) - \frac{k_q R_0}{\sigma_\mu}(e^{\sigma_\mu (t-\tau_D)} - 1) \times u(t) \\
Q_c(t) &= \frac{k_q R_0}{\sigma_\mu + k_d}(e^{(\sigma_\mu + k_d)(t-\tau_D)} e^{k_d \tau_D} - e^{k_d \tau_D}) e^{-k_d t} \times u(t)
\end{aligned}
$$

and

$$u(t) = 0$$

for

$$t < \tau_D$$

We have measurements of the total cell volume, i.e. $R(t) + Q_a(t) + Q_c(t)$ at several time points, for cell cultures with the following structure:

- 8 genetic designs, comprising 7 genetic interventions and a control design.
- Between 3 and 4 clones implementing each design.
- Two technical replicates for each clone.
- For each technical replicate, measurements of total cell volume at 24 hour intervals for between 3 and 5 days.

Replicates of the same clone are taken to be biologically identical, though we expect some variation in measurements due to the experimental conditions. Clones with the same design are expected to be similar, with some degree of clonal variation that may differ depending on the design. There is no prior information distinguishing the designs from each other, or distinguishing clones with the same design.

Given these assumptions a multi-level Bayesian statistical model is appropriate.

Measurement model

We used the following measurement model:

$$y \sim log\, normal(\log(\hat{y}(t, R_0, \mu, \tau, k_q, k_d)), \sigma)$$

where

- R_0, τ_D, k_q and k_d are vectors of clone-level parameters.
- μ is a model parameter representing the pre-treatment growth rate, which we assume is the same for all replicates.
- σ is an unknown log-scale error standard deviation.
- t is a vector of known measurement times (one per measurement).
- $\hat{y}$ is a function mapping parameter configurations to densities, under the delay differential equation assumptions set out above.

We believe that the measurement error will be proportional to the true viable cell density for most measurements, motivating the use of the log-normal generalised linear model. However, we hypothesise that for cell densities below 0.3×10^{-6} viable cells per mL this will not be the case, as for these measurements the error is dominated by factors that do not depend on the true cell density, such as distortion due to fragmentation of dead cells.

To allow the model to incorporate this postulated effect we use the following distributional model:

$$\sigma = exp(a_\sigma + b_\sigma * \min(0, \ln(\hat{y} - 0.3)))$$

In this equation the parameter b_σ represents the degree to which the log-scale measurement error increases or decreases as the true value gets lower than 0.3. Figure 1 below shows the effect of various values of b_σ on the ln-scale error standard deviation.

True density vs ln scale measurement standard deviation for a range of b_σ values

Code to generate figure 1:

```
import numpy as np
from matplotlib import pyplot as plt

plt.style.use("sparse.mplstyle")
y = np.linspace(0.04, 0.4, 20)
bs = [-0, -0.05, -0.1, -0.15, -0.2]
diff = np.array([np.log(yi/0.3) if yi < 0.3 else 0 for yi in y])
for b in bs:
    plt.plot(y, 0.2 + b * diff, label=str(b))
plt.legend(title="$b_\sigma$", frameon=False)
plt.xlabel("True density")
plt.xlabel("ln scale standard deviation")
plt.savefig("results/plots/y_vs_log_sd.png", bbox_inches="tight")
```

Parameters

The clone-level vectors τ_D and k_d are treated as determined by other parameters as follows:

$$\ln(\tau) = \tau const + d_\tau + c_{tau}$$
$$\ln(k_d) = dconst + d_d + c_d$$

In these equations

- $qconst$, $\tau const$ and $dconst$ are single unknown numbers representing the (log scale) mean parameter values with no interventions

- X is a matrix indicating which clones have which interventions
- d_τ and d_d are vectors of unknown intervention effects, with the parameter for the empty intervention fixed at zero.
- c_τ and c_d are vectors of unknown clone effects, representing random clonal variation.

To investigate whether the genetic interventions measurably affected the rate at which cells transition from the normal state R to the growth arrest state Q_a, we compared two different ways of modelling the clone-level vectors k_q. In the first model design M1, k_q is treated like τ_D and k_d, i.e

$$\begin{aligned} \ln k_q &= qconst + d_q + c_q \\ d_q &\sim N(0, 0.3) \end{aligned}$$

In the second design M2, we assume that there are no design-level effects, i.e.

$$\ln k_q = qconst + c_q$$

To represent our scientific knowledge we used informative log-normal prior distributions for τ_D, k_d and k_q based on quantiles.

Parameter	1% Quantile	99% Quantile
τ_D	0.4	7.5
k_q	1	5
k_d	0.05	2.5

The priors for the design effect parameters d_τ and d_d were as follows:

$$\begin{aligned} d_\tau &\sim N(0, 0.3) \\ d_d &\sim N(0, 0.3) \end{aligned}$$

The clonal variation parameters c_τ, c_q and c_d have independent normal prior distributions:

$$\begin{aligned} c_\tau &\sim normal(\mathbf{0}, 0.1) \\ c_q &\sim normal(\mathbf{0}, 0.1) \\ c_d &\sim normal(\mathbf{0}, 0.1) \end{aligned}$$

Other parameters have informative prior distributions based on scientific knowledge:

Parameter	Distribution	1% Quantile	99% Quantile
μ	log normal	0.65	0.73
R_0	log normal	2	3
a_σ	log normal	0.05	0.13
b_{sigma}	normal	-0.3	0.3

Data representation

We believed that there should be no noticeable effect due to the Δbok intervention compared to the *Empty plasmid* control. To verify that this was the case we compared the results of fitting the models M1 and M2 with and without distinguishing the Δbok design from the other designs.

Model comparison

To compare different models we calculated the approximate leave-one-timecourse-out log predictive density for models M1 and M2 on puromycin and sodium butyrate treatments using the python package arviz (v0.11.2). See (Vehtari, Gelman, and Gabry 2017) for more about this model comparison method. As some timecourses had a Pareto-k diagnostic higher than 0.7, indicating that the approximation was inaccurate, we refitted models in order to find the exact leave-one-out log predictive density.

We further evaluated our models using graphical posterior predictive checks and by inspecting the modelled parameter values.

Results

Model comparison

We tested four non-null model designs on four treatments, finding their estimated log predictive density, or elpd, using the semi-approximate leave-one-timecourse-out process described above. The results are exemplified for puromycin treatment, and are presented in the following sections.

To verify that the design effects were informative, we also fit a null model with no design effects. This model performed markedly worse according to our semi-approximate elpd test.

The models are listed in order of elpd. The column Δ elpd shows the estimated difference each model's elpd and that of the best model. The column SE Δelpd shows the estimated standard error of this difference.

RELOO Comparison results: Puromycin data

Model	elpd	SE epld	Δelpd	SE Δelpd
m2_ab	-22.132365	25.259048	0.000000	0.000000
m2_abc	-23.403847	27.501952	1.271481	7.795202
m1_abc	-32.594331	35.229694	10.46196	14.34333
m1_ab	-33.396555	27.556028	11.26419	4.564483
null	-92.364474	29.310052	70.23210	17.37136

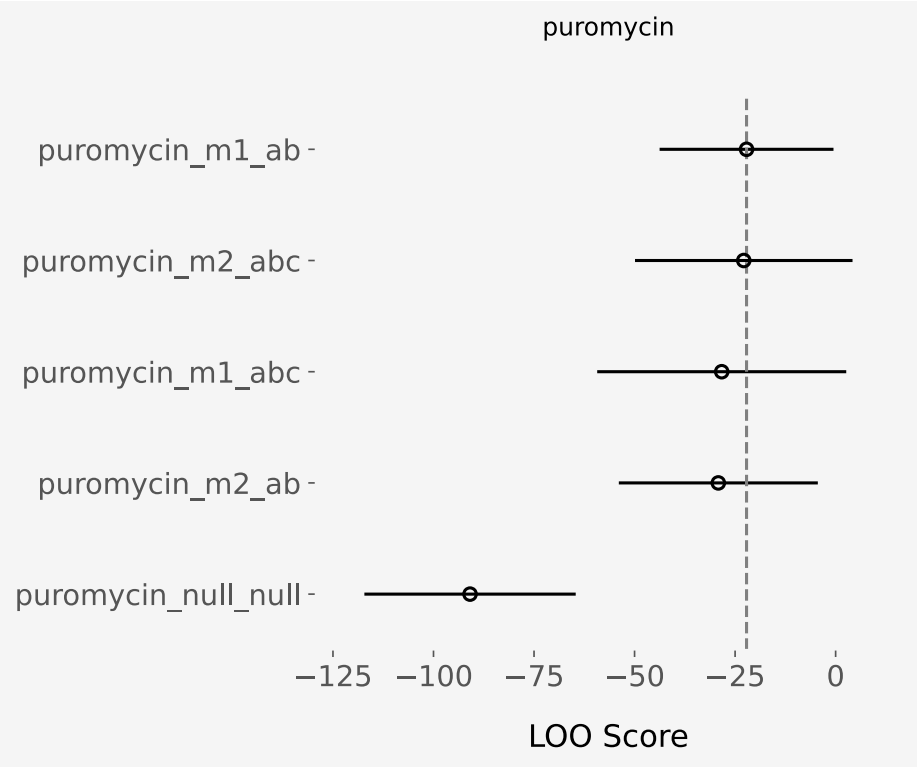


Figure 1: Puromycin treatment RELOO comparison

We noted that the simplest and easiest to interpret model design (`m2_ab`) scored highest in leave-one-timecourse-out elpd, and we therefore chose to use this simpler model knowing that doing so did not entail a tangible sacrifice of predictive power.

Observed vs modelled timecourses

The figures in this section show observed timecourses for each design under the 15 ug/mL Puromycin treatment alongside 99% posterior predictive intervals, for each model and treatment.

The modelled and observed timecourses appear qualitatively similar in all cases, suggesting that none of the models are dramatically mis-specified.

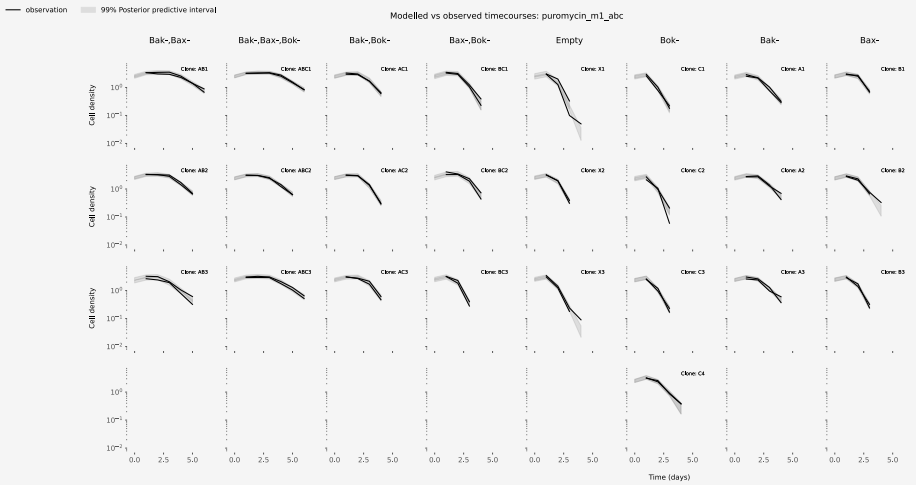


Figure 2: Puromycin treatment, design m1_abc

Posterior distributions of design parameters

The figures below plot the 2.5% to 97.5% marginal posterior intervals for the log-scale design-level effects, relative to the control experiment.

Some models are able to detect clear design effects with respect to the τ and k_d parameters. For example, for model `m2_ab` the posterior for the effect of design B on the delay parameter τ_d concentrates above zero.

On the other hand, all of the posterior intervals for design-specific effects on the parameter k_q include zero, showing that none of the models can detect any impact from the designs on the speed of transition to growth arrest.

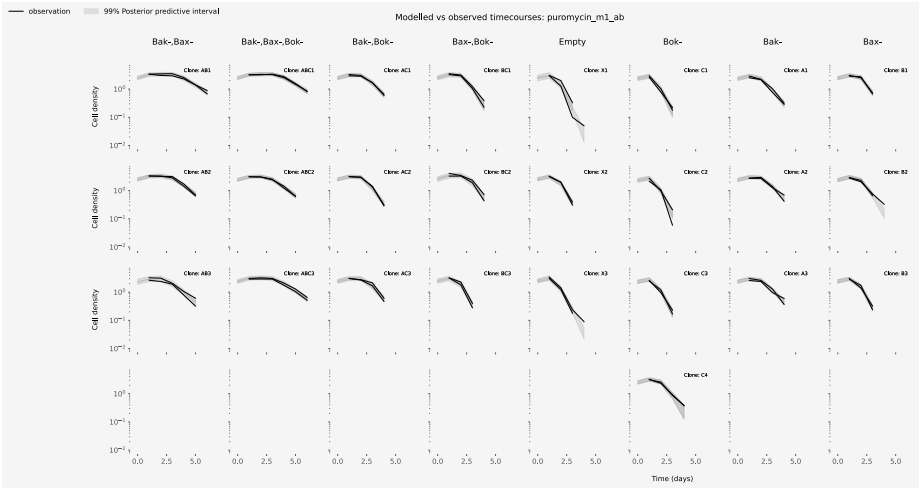


Figure 3: Puromycin treatment, design m1_ab

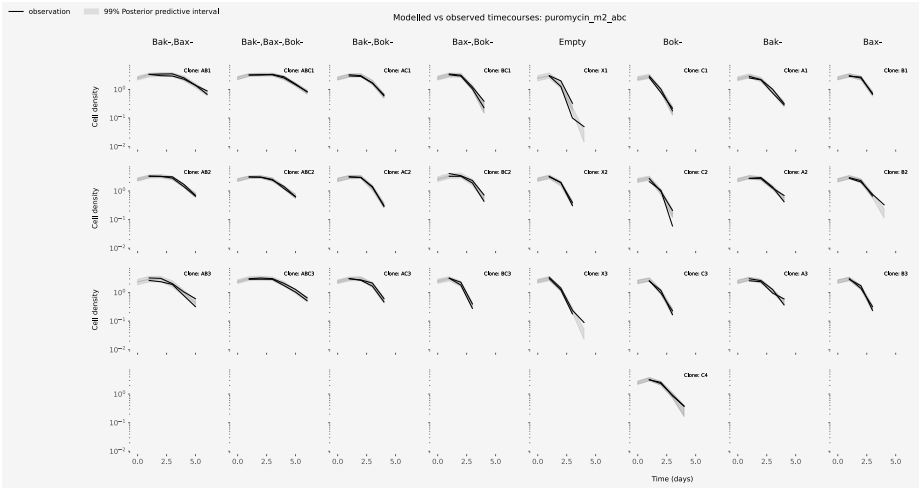


Figure 4: Puromycin treatment, design m2_abc

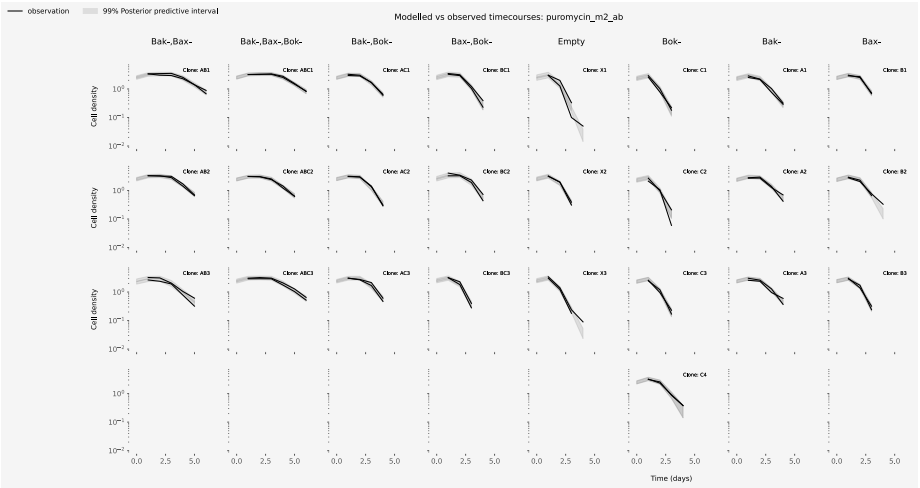
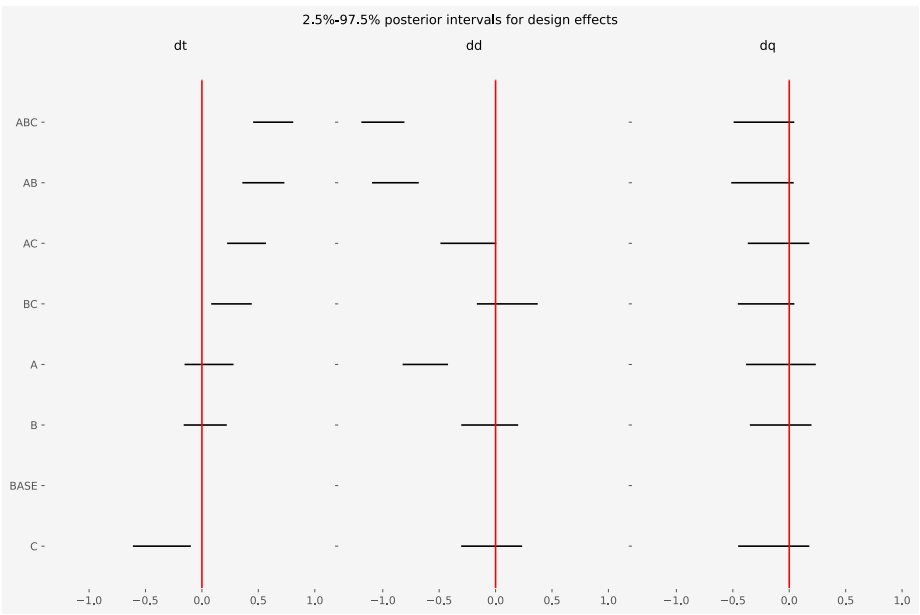
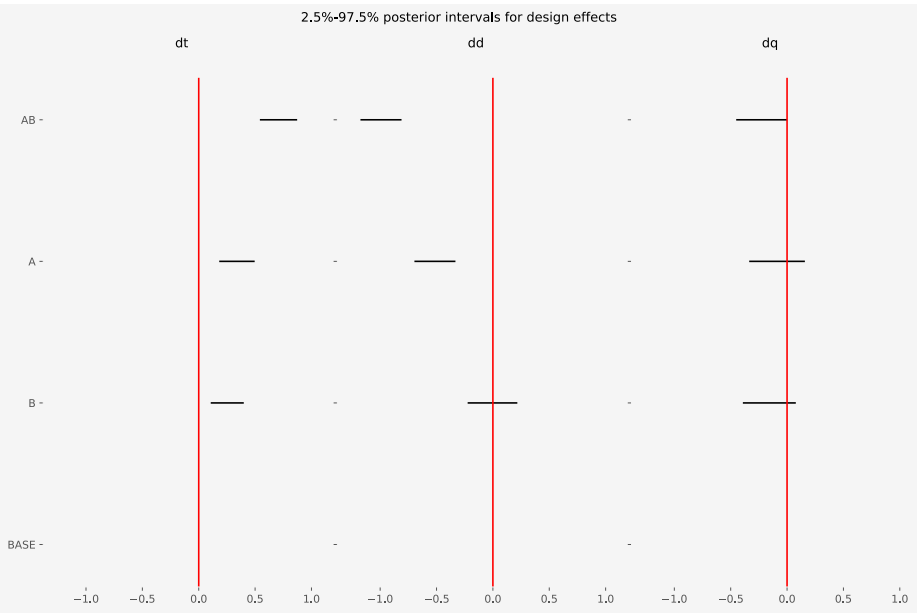


Figure 5: Puromycin treatment, design m2_ab

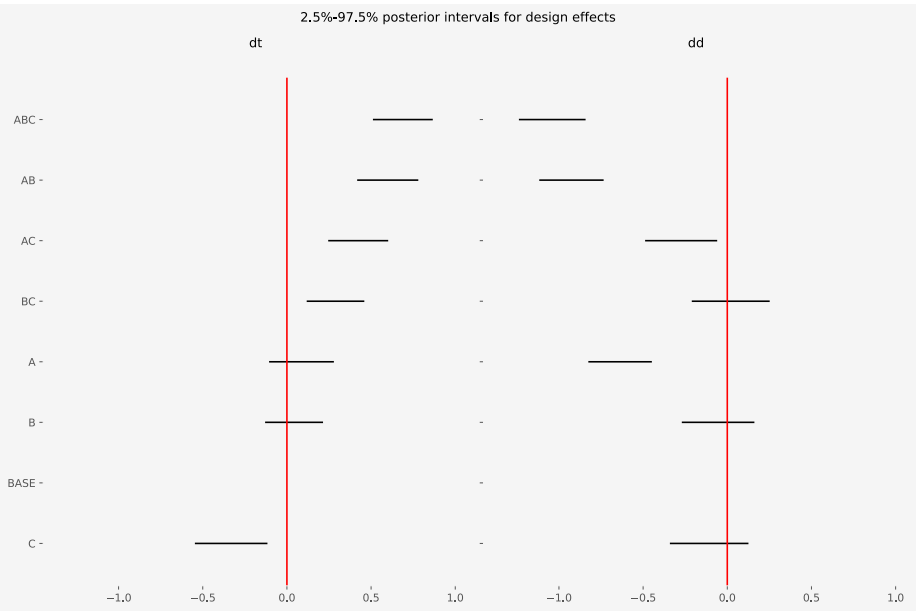
Posterior intervals for design level parameters, treatment 15ug/mL Puromycin, design m1_abc



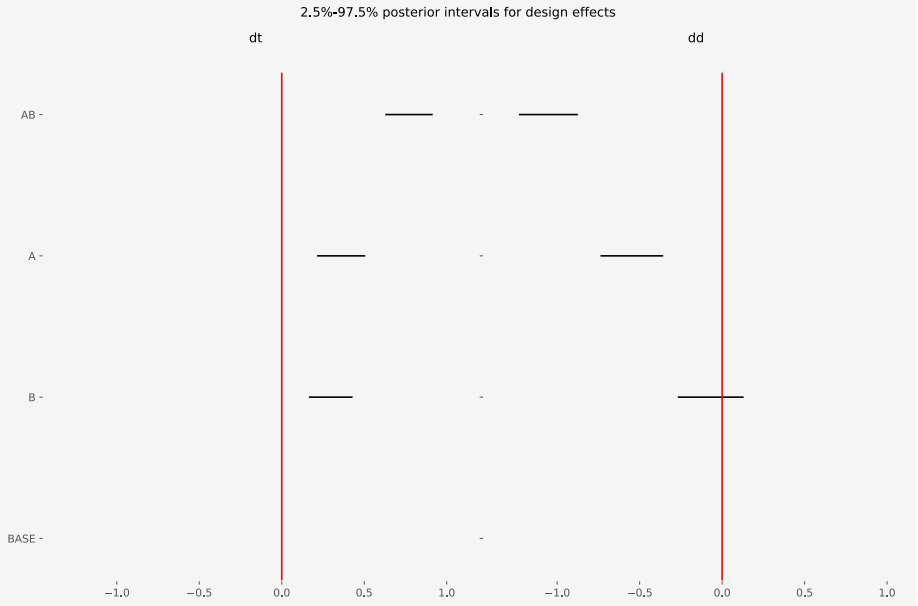
Posterior intervals for design level parameters, treatment 15ug/mL Puromycin, design m1_ab



Posterior intervals for design level parameters, treatment 15ug/mL Puromycin, design m2_abc



Posterior intervals for design level parameters, treatment 15ug/mL Puromycin, design m2_ab



References

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2017. "Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC." *Statistics and Computing* 27 (5): 1413–32. https://doi.org/10.1007/s11222-016-9696-4.